arXiv:cond-mat/0403030v2 [cond-mat.stat-mech] 31 Mar 2005

A Statistical Mechanics Model for the Emergence of Consensus

Giacomo Raffaelli and Matteo Marsili
INFM-SISSA, via Beirut 2-4, Trieste I-34014, Italy and
Abdus Salam International Centre for Theoretical Physics
Strada Costiera 11, 34014 Trieste Italy

The statistical properties of pairwise majority voting over S alternatives is analyzed in an infinite random population. We first compute the probability that the majority is transitive (i.e. that if it prefers A to B to C, then it prefers A to C) and then study the case of an interacting population. This is described by a constrained multi-component random field Ising model whose ferromagnetic phase describes the emergence of a strong transitive majority. We derive the phase diagram, which is characterized by a tri-critical point and show that, contrary to intuition, it may be more likely for an interacting population to reach consensus on a number S of alternatives when S increases. This effect is due to the constraint imposed by transitivity on voting behavior. Indeed if agents are allowed to express non transitive votes, the agents' interaction may decrease considerably the probability of a transitive majority.

I. INTRODUCTION

Social choice and voting theory address the generic problem of how the individual preferences of N agents over a number S of alternatives can be aggregated into a social preference. This issue involves collective phenomena, such as the emergence of a common opinion in a large population, which have attracted some interest in statistical physics. For example, the voter [1] and random field Ising models [2] have been proposed to study how the vote's outcome between two alternatives is affected when voters influence each other. In the case of two alternatives ($S = 2$) the statistical mechanics of the majority vote model has also been numerically studied on random graphs [3]. In general, the framework for this kind of studies is the statistical mechanics approach to socio-economic behavior [4, 5], which stems from realizing that the emergence of "macro-behavior" can be the result of the interaction of many agents, each with their own beliefs and expectations.

The majority rule can be naturally extended to $S > 2$ alternatives by considering the social preferences stemming from majority voting on any pair of alternatives, i.e. pairwise majority rule (PMR). This extension however is problematic, as observed back in 1785 by Marquis de Condorcet [6]. He observed that the PMR among three alternatives may exhibit an irrational behavior, with the majority preferring alternative A to B, B to C and C to A, even though each individual has transitive preferences. These so-called Condorcet's cycles may result in the impossibility to determine a socially preferred alternative or a complete ranking of the alternatives by pairwise majority voting (see also Ref. [7] for a relation with statistical mechanics of dynamical systems). PMR is not the only way to aggregate individual rankings into a social preference [8, 9]. However the situation does not improve much considering other rules. For example, the transitivity of social preferences is recovered by resorting to voting rules like Borda count, where each voter assigns a score to each alternative, with high scores corresponding to preferred alternatives. It turns out that these rules also violate some other basic requirement. The basic desiderata of a social choice rule are that it should be able to rank all alternatives for whatever individual preferences (*unrestricted domain*), it should be *transitive*, it should be *monotonous*, i.e. the social rank of an alternative A cannot decrease when an individual promotes A to a higher rank, and it should be *independent of irrelevant alternatives*, i.e. the social preference between A and B cannot depend on the preferences for other alternatives (independence of irrelevant alternatives is important because it rules out the possibility of manipulating the election's outcome by falsely reporting individual preferences). For example, in plurality voting each individual casts one vote for his top candidate and candidates are ranked according to the number of votes they receive. This satisfies all requirement but the last one, as vividly illustrated by recent election outcomes [8].

The discomfort of social scientists with the impossibility to find a reasonable voting rule has been formalized by Arrow's celebrated theorem [9]. This states that a social choice rule that satisfies all of the above requirements has to be *dictatorial*, that is there exists an agent – the dictator – such that the social preference between any two alternatives is the preference of that agent.

A way to circumvent the impasse of this result is to study the properties of social choice rules on a restricted domain of possible individual preferences. For example in politics, it may be reasonable to rank all candidates from extreme left to extreme right. If the preferences of each individual has a "single peak" when candidates are ranked in this order (or any other order), then pairwise majority is transitive. It has recently been shown that pairwise majority turns out to be the rule which satisfies all requirements in the largest domain [8], thus suggesting that pairwise majority is the best possible social choice rule.

In this paper we first try to quantify how good is majority rule by estimating the probability that pairwise majority yields a transitive preference relation in a typical case where individual preferences are drawn at ran-

dom. This and closely related issues have been addressed by several authors [10–12].

Secondly, we study how the situation changes when agents influence each other. In particular, as in the $S = 2$ [1, 2], we restrict to the relevant case where the interaction arises from conformism [5]. Basically conformism can stem from three different reasons [15]. It can be *pure* or *imitative*, because people simply want to be like others. It can be due to the fact that in some cases conforming facilitates life (*instrumental conformism*). Or it can be due to people deriving information about the value of a choice from other people's behavior (*informational conformism*). In this light, our results may shed light on a number of social phenomena, ranging from fashions or fanaticism, where conformism may lead to the rise and spread of broadly accepted systems of values, to the questions of how much information should the agents share in order to achieve consensus on S items. At any rate, our discussion will focus on the consequence of conformism on the collective behavior, without entering into details as to where this conformism stems from.

We show that the occurrence of a transitive social choice on a number S of alternatives for any choice of the individual preferences, is related to the emergence of spontaneous magnetization in a multi-component Ising model. We find a phase diagram similar to that of the single component model [14] with a ferromagnetic phase and a tricritical point separating a line of second order phase transitions from a first order one. The ferromagnetic state describes the convergence of a population to a common and transitive preference ranking of alternatives, due to social interaction.

Remarkably, we find that the ferromagnetic region expands as S increases. Hence while without interaction the probability $P(S)$ of a transitive majority vanishes rapidly as S increases, if the interaction strength is large enough, the probability of a transitive majority increases with S and it reaches one for S large enough. In other words, an interacting population may reach more likely consensus when the complexity of the choice problem (S) increases.

We finally contrast these findings with the case where agents need not express a transitive vote (e.g. they may vote for A when pitted against B, for B agains C and for C against A). This is useful because we find that then the probability of finding a transitive majority is much lower. In other words, individual coherence is crucial for conformism to enforce a transitive social choice.

II. NON-INTERACTING POPULATION

We shall first describe the behavior of a non-interacting population and then move on to the interacting case. Let us consider a population of N individuals with preferences over a set of S choices or candidates. We shall mainly be interested in the limit $N \to \infty$ of an infinite population. We limit attention to strict preferences, i.e. we rule out the case where agents are indifferent between items. Hence preference relations are equivalent to rankings of the S alternatives. It is convenient to represent rankings with matrices $\hat{\Delta}_i$ for each agent $i = 1, \ldots, N$, whose elements take values $\Delta_i^{ab} = +1$ or -1 if i prefers choice a to $b \neq a$ or vice-versa, with $a, b = 1, \ldots, S$. Notice that $\Delta_i^{ba} = -\Delta_i^{ab}$. Let $\mathcal{R}$ be the set of matrices $\hat{\Delta}$ which correspond to a transitive preference relation. Clearly the number of such matrices equals the number $|\mathcal{R}| = S!$ of rankings of the S alternatives. Hence not all the $2^{S(S-1)/2}$ possible asymmetric matrices with binary elements $\Delta_i^{ab} = \pm 1$ correspond to acceptable preference relations. For example, if $\Delta^{1,2} = \Delta^{2,3} = \Delta^{3,1}$ then $\hat{\Delta} \notin \mathcal{R}$. We use the term ranking to refer to matrices $\hat{\Delta} \in \mathcal{R}$ in order to avoid confusion later, when we will introduce preferences over rankings, i.e. over elements of $\mathcal{R}$. We assume that each agent i is assigned a ranking $\hat{\Delta}_i$ drawn independently at random from $\mathcal{R}$.

In order to compute the probability $P(S)$ that pairwise majority yields a transitive preference relation, in the limit $N \to \infty$, let us introduce the matrix $\hat{x} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \hat{\Delta}_i$. The assumption on $\hat{\Delta}_i$ implies that the distribution of x^{ab} is Gaussian for $N \to \infty$ and it is hence completely specified by the first two moments $\langle x^{ab} \rangle = 0$ and $\langle x^{ab} x^{cd} \rangle = \{\mathcal{G}^{-1}\}^{ab,cd}$ which is 0 except for $\{\mathcal{G}^{-1}\}^{ab,ab} = 1$, $\{\mathcal{G}^{-1}\}^{ab,ad} = \{\mathcal{G}^{-1}\}^{ab,cb} = 1/3$ and $\{\mathcal{G}^{-1}\}^{ab,ca} = \{\mathcal{G}^{-1}\}^{ab,bd} = -1/3$, where we have introduced the notation $\mathcal{M}$ for matrices with elements $M^{ab,cd}$. The matrix $\mathcal{G}^{-1}$ can be inverted by a direct computation, and we find that the matrix $\mathcal{G}$ has the same structure of $\mathcal{G}^{-1}$ but with $\mathcal{G}^{ab,ab} = 3\frac{S-1}{S+1}$, $\mathcal{G}^{ab,ad} = -\frac{3}{S+1} = \mathcal{G}^{ab,cb} = -\mathcal{G}^{ab,bd} = -\mathcal{G}^{ab,ca}$.

Let us first compute the probability $CW(S)$ that one of the alternatives, is better than all the others. This means that there is a consensus over the winner, while nothing is assumed for the relations between the other choices. The preferred alternative is known in social choice literature as the Condorcet winner, and much interest has been devoted to it, since the presence of such a preferred alternative saves at least the possibility of electing a favorite choice. $CW(S)$ is just the probability that $x^{1,a} > 0$ for all $a > 1$ multiplied by S. In this way, we recover a known results [11], which can be conveniently casted in the form

$$CW(S) = S\sqrt{\frac{2}{\pi}} \int_{-\infty}^{\infty} e^{-2y^2 + (S-1)\log[\mathrm{erfc}(y)/2]}. \qquad (1)$$

Notice that $CW(S)$ is much larger than the naïve guess $S/2^{S-1}$, derived assuming that $x^{ab} > 0$ occurs with probability $1/2$ for all ab. Indeed asymptotic expansion of Eq. (1) shows that

$$CW(S) \simeq \sqrt{\frac{\pi}{2}} \frac{\sqrt{\log S}}{S} \left[1 + O(1/\sqrt{\log S})\right]$$

decays extremely slowly for $S \gg 1$.

The probability that the majority ranking is equal to the cardinal one ($1 \succ 2 \succ \ldots \succ S$) is given by the prob-

ability that $x^{ab} > 0$ for all $a < b$. This is only one of the $S!$ possible orderings, then the probability of a transitive majority can be written as

$$P(S) = S! \frac{[3/(2\pi)]^{\frac{S(S-1)}{4}}}{(S+1)^{\frac{S-1}{2}}} \int_0^\infty d\hat{x} \exp\left[-\frac{1}{2}\hat{x}\cdot\mathcal{G}\cdot\hat{x}\right] \quad (2)$$

where $\int_0^\infty d\hat{x} \equiv \int_0^\infty dx_{1,2}\ldots\int_0^\infty dx_{S-1,S}$ and we defined the product $\hat{r}\cdot\hat{q} = \sum_{a<b} r^{ab}q^{ab}$ and its generalization to matrices $\hat{r}\cdot\mathcal{M}\cdot\hat{q} = \sum_{a<b}\sum_{c<d} r^{ab}M^{ab,cd}q^{cd}$. The normalization factor is computed from the spectral analysis of $\mathcal{G}$ [20].

We were not able to find a simpler form for this probability. Fig. 1 reports Montecarlo estimates of $P(S)$. For $S = 3$ we recover the result [10]

$$P(3) = CW(3) = \frac{3}{4} + \frac{3}{2\pi}\sin^{-1}\frac{1}{3} \cong 0.91226\ldots \quad (3)$$

Again the naïve guess $P(S) \approx S!/2^{S(S-1)/2}$ based on the fraction of acceptable rankings largely underestimates this probability. This means that the collective behavior of the majority hinges upon the (microscopic) transitivity of individual rankings.

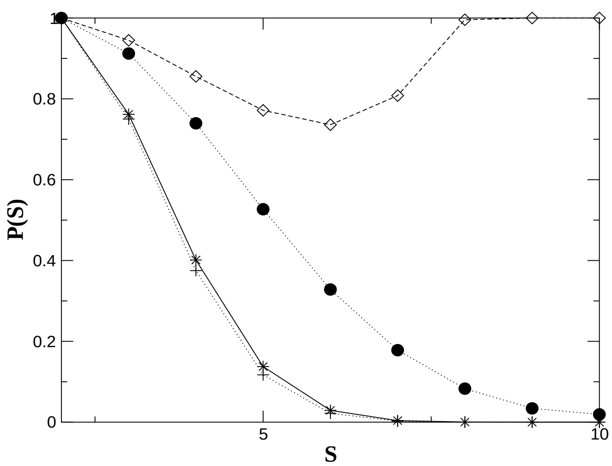

FIG. 1: Probability $P(S)$ of a transitive majority (•) compared to the naïve guess $S!/2^{S(S-1)/2}$ (+). (◇) shows the case of an interacting population with $\beta = 0.45$ and $\epsilon = 0.8$, see Section III, $*$ show the same case for the unconstrained case.

III. INTERACTING VOTERS

Let us now introduce interaction among voters. We assume that agents have an *a-priori* transitive preference over the alternatives, specified by a ranking $\hat{\Delta}_i \in \mathcal{R}$. We allow however agents to have a voting behavior which does not necessarily reflect their *a-priori* ranking, that is, we introduce a new matrix $\hat{v}_i$ such that $v_i^{ab} = +1$ (-1) if agent i, in a context between a and b, votes for a (b). We will first study the case when $\hat{v}_i \in \mathcal{R}$, which corresponds to agents having a rational voting behavior. This means that even though an agent is influenced by others, she will maintain a coherent choice behavior (transitivity). We will contrast this case with that where the constraint on individual coherence $\hat{v}_i \in \mathcal{R}$ is removed.

To account for interaction, the matrix $\hat{v}_i$ depends not only on agents' preferences $\hat{\Delta}_i$, but also on the interaction with other agents. Within economic literature, this dependence is usually introduced by means of an utility function u_i which agents tend to maximize. Notice that this utility function represents a preference over preferences (rankings).

Formally, this utility function depends both on an idiosyncratic term $\hat{\Delta}_i \in \mathcal{R}$ describing the *a priori* ranking, and on the behavior of other agents, $\hat{v}_{-i} \equiv \{\hat{v}_j,\ \forall j \neq i\}$, through the majority matrix

$$\hat{m} = \frac{1}{N}\sum_{i=1}^{N}\hat{v}_i. \quad (4)$$

More precisely, we define an utility function

$$u_i(\hat{v}_i, \hat{v}_{-i}) = (1-\epsilon)\hat{\Delta}_i\cdot\hat{v}_i + \epsilon\hat{m}\cdot\hat{v}_i. \quad (5)$$

where the last term captures conformism as a diffuse preference for aligning to the majority [5, 15]. For $\epsilon = 0$ maximal utility in Eq. (5) is attained when agents vote as prescribed by their *a priori* rankings, i.e. $\hat{v}_i = \hat{\Delta}_i$ $\forall i$. On the contrary, for $\epsilon = 1$ agents totally disregard their rankings and align on the same ranking $\hat{v}_i = \hat{m}$ $\forall i$, which can be any of the $S!$ possible ones.

Let us characterize the possible stable states, i.e. the Nash equilibria of the game defined by the payoffs of Eq. (5). These are states $\hat{v}_i^*$ such that each agent has no incentives to change his behavior, if others stick to theirs, i.e. $u_i(v_i, v_{-i}^*) \leq u_i(v_i^*, v_{-i}^*)$ for all i. The *random* state $\hat{v}_i^* = \hat{\Delta}_i$ is (almost surely) a Nash equilibrium $\forall\epsilon < 1$, because the payoff of aligning to the majority $\hat{m} = \hat{x}/\sqrt{N}$ is negligible with respect to that of voting according to own ranking $\hat{\Delta}_i$. Then we have $u_i(\hat{\Delta}_i, \hat{\Delta}_{-i}) = \frac{S(S-1)}{2}[1-\epsilon+\epsilon O(1/\sqrt{N})]$. This Nash equilibrium is characterized by a majority which is not necessarily transitive, i.e. which is transitive with probability $P(S) < 1$ for $N \gg 1$.

Also *polarized* states with $\hat{v}_i = \hat{m}$ for all i are Nash equilibria for $\epsilon > 1/2$. Indeed, with some abuse of notation, when all agents take $\hat{v}_j = \hat{m}$ for some $\hat{m}$, agent i receives an utility $u_i(\hat{m}, \hat{m}) = (1-\epsilon)\hat{\Delta}_i\cdot\hat{m} + \frac{S(S-1)}{2}\epsilon$. The agents who are worse off are those with $\hat{\Delta}_i = -\hat{m}$ for whom $u_i(\hat{m}, \hat{m}) = \frac{S(S-1)}{2}[2\epsilon - 1] \cong -u_i(\hat{\Delta}_i, \hat{m}) + O(1/N)$. Then as long as $\epsilon > 1/2$, even agents with $\hat{\Delta}_i = -\hat{m}$ will not profit from abandoning the majority. Therefore $\hat{v}_i = \hat{m}$ for all i is a Nash equilibrium. Notice that whether the majority is transitive ($\hat{m} \in \mathcal{R}$) or not depends on whether agents express transitive preferences ($\hat{v}_i \in \mathcal{R}$) or not. In the former case the majority will be transitive whereas if non transitive voting is allowed

there is no need to have $\hat{m} \in \mathcal{R}$ and there are $2^{S(S-1)/2}$ possible polarized Nash equilibria. Only in $S!$ of them the majority is transitive (i.e. when $\hat{m} \in \mathcal{R}$).

It is easy to check that there are no other Nash equilibria. Summarizing, for $\epsilon > 1/2$ there are many Nash equilibria. Depending on the dynamics by which agents adjust their voting behavior one or the other of these states will be selected.

IV. STATISTICAL MECHANICS OF INTERACTING VOTERS

Strict utility maximization leads to the presence of multiple equilibria, leaving open the issue of which equilibrium will the population select. It is useful to generalize the strict utility maximization into a stochastic choice behavior which allows for mistakes (or experimentation) with a certain probability [16]. This on one side may be realistic in modelling many socio-economic phenomena [4, 15, 17]. On the other this rescues the uniqueness of the solution, in terms of the probability of occurrence of a given state $\{\hat{v}_i\}$, under some ergodicity hypothesis. Here, as in [17], we assume that agents have the following probabilistic choice behavior: agents are asynchronously given the possibility to revise their voting behavior. When agent i has a revision opportunity, he picks a voting profile $\hat{w}$ ($\in \mathcal{R}$ when voters are rational) with probability

$$P\{\hat{v}_i = \hat{w}\} = Z_i^{-1} e^{\beta u_i(\hat{w}, \hat{v}_{-i})} \tag{6}$$

where Z_i is a normalization constant. Without entering into details, for which we refer to Ref. [17], let us mention that Eq. (6) does not necessarily assume that agents randomize their behavior on purpose. It models also cases where agents maximize a random utility with a deterministic term u_i and a random component. Then the parameter β is related to the degree of uncertainty (of the modeler) on the utility function [18].

When agent i revises his choice the utility difference $\delta u_i = u_i(\hat{v}_i, \hat{v}_{-i}) - u_i(\hat{v}'_i, \hat{v}_{-i})$ for a change $\hat{v}_i \to \hat{v}'_i$ is equal to the corresponding difference in $-H$, where

$$H\{\hat{v}_i\} = -(1-\epsilon)\sum_{i=1}^{N} \hat{\Delta}_i \cdot \hat{v}_i - \frac{\epsilon}{2N}\sum_{i,j=1}^{N} \hat{v}_j \cdot \hat{v}_i. \tag{7}$$

hence in the long run, the state of the population will be described by the Gibbs measure $e^{-\beta H}$ because the dynamics based on Eq. (6) satisfies detailed balance with the Gibbs measure.

H in Eq. (7) is the Hamiltonian of a multi-component random field Ising model (RFIM) where each component v_i^{ab} with $a < b$ is a component of the spin, $\hat{\Delta}_i$ represents the random field and the term $\frac{\epsilon}{2N}\sum_{i,j=1}^{N} \hat{v}_j \cdot \hat{v}_i$ is a mean field interaction. Indeed $\hat{v}_i$ has $S(S-1)/2$ components which take values $v_i^{ab} = \pm 1$. The peculiarity of this model is that the components of the fields $\hat{\Delta}$ are not independent. Indeed not all the $2^{S(S-1)/2}$ values of $\hat{\Delta}_i$ are possible but only those $\hat{\Delta}_i \in \mathcal{R}$, which are $S!$. The same applies to the spin components $\hat{v}_i$ when rational voting behavior is imposed. Were it not for this constraint, the model would just correspond to a collection of $S(S-1)/2$ uncoupled RFIM.

The statistical mechanics approach of the RFIM [13, 14] can easily be generalized to the present case. The partition function can be written as

$$Z(\beta) = \mathrm{Tr}_{\{\hat{v}_i\}} e^{-\beta H} = \int d\hat{m} e^{-N\beta f(\hat{m})} \tag{8}$$

where the trace $\mathrm{Tr}_{\{\hat{v}_i\}}$ over spins runs on all $\hat{v}_i \in \mathcal{R}$ when voting behavior is rational, or over all $\hat{v}_i$ otherwise. The free energy $f(\hat{m})$ is given by

$$f(\hat{m}) = \frac{\epsilon}{2}\hat{m}^2 - \frac{1}{N\beta}\sum_{i=1}^{N} \log\left[\sum_{\hat{v}} e^{\beta[(1-\epsilon)\hat{\Delta}_i + \epsilon\hat{m}]\cdot\hat{v}}\right] \tag{9}$$

where once again the sum over the $\hat{v}_i$ runs inside $\mathcal{R}$ if agents are rational, or is not limited otherwise. It is evident that f is self averaging. Hence in the limit $N \to \infty$ we can replace $\frac{1}{N}\sum_i \ldots$ with the expected value $\frac{1}{S!}\sum_{\hat{\Delta}\in\mathcal{R}} \ldots \equiv \langle \ldots \rangle_\Delta$ on $\hat{\Delta}_i$. It is also clear that the integral over $\hat{m}$ of Eq. (8) in this limit is dominated by the saddle point value

$$\hat{m} = \left\langle \frac{\sum_{\hat{v}} \hat{v} e^{\beta[(1-\epsilon)\hat{\Delta} + \epsilon\hat{m}]\cdot\hat{v}}}{\sum_{\hat{v}} e^{\beta[(1-\epsilon)\hat{\Delta} + \epsilon\hat{m}]\cdot\hat{v}}} \right\rangle_\Delta \tag{10}$$

This equation can be solved from direct iteration and shows that for large enough values of $\beta > \beta_c$ there is a transition, as ϵ increases, from a paramagnetic state with $\hat{m} = 0$ to a polarized (ferromagnetic) state where $\hat{m} \neq 0$. In Fig. (2) we plot the result of such iterative solution for the RFIM case, i.e. when $S = 2$. Since for some values T, ϵ both the ferromagnetic and the paramagnetic state can be stable, we have solved for the magnetization $\hat{m}$ starting both from a $\hat{m} = 0$ and from $\hat{m} = \hat{1}$ states. Then we selected the correct equilibrium state by comparing the free energy of the different solutions. The stability of the paramagnetic solution $\hat{m} = 0$ can be inferred from the expansion of Eq. (10) around $\hat{m} = 0$, which reads

$$\hat{m} = \beta\epsilon\mathcal{J}\cdot\hat{m} + O(\hat{m}^3) \tag{11}$$

where

$$J^{ab,cd} = \left\langle \left\langle v^{ab} v^{c,d} | \hat{\Delta} \right\rangle_v - \left\langle v^{ab} | \hat{\Delta} \right\rangle_v \left\langle v^{c,d} | \hat{\Delta} \right\rangle_v \right\rangle_\Delta . \tag{12}$$

Here averages $\left\langle \ldots | \hat{\Delta} \right\rangle_v$ over $\hat{v}$ are taken with the distribution

$$P(\hat{v}|\hat{\Delta}) = \frac{e^{\beta(1-\epsilon)\hat{\Delta}\cdot\hat{v}}}{\sum_{\hat{u}} e^{\beta(1-\epsilon)\hat{\Delta}\cdot\hat{u}}}. \tag{13}$$

When the largest eigenvalue Λ of $\beta\epsilon\mathcal{J}$ is larger than one, the paramagnetic solution $\hat{m} = 0$ is unstable and only the polarized solution $\hat{m} \neq 0$ is possible. In Fig.2 the line that marks the region of instability of the paramagnetic solution is plotted at the bottom.

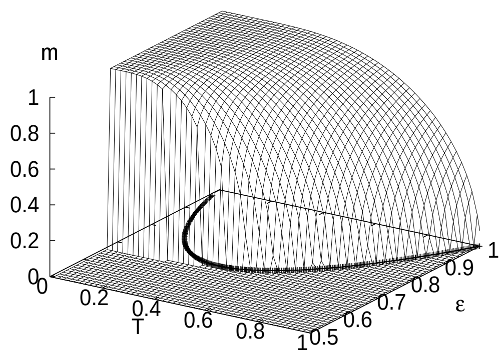

FIG. 2: Phase diagram for $S = 2$. The plot shows the magnetization, while on the bottom we have drawn the line that marks the instability of the paramagnetic solution.

A. Constrained case, $\hat{v}_i \in \mathcal{R}$

Here both the individual *a-priori* rankings $\hat{\Delta}_i$ and the voting behavior $\hat{v}_i$ of each agent are transitive. Results for the numerical iteration of Eq. (10) are shown in the inset of Fig.3 for different values of β and for $S = 5$. Fig. 3 shows the phase diagram for $S = 2,\ 3$ and for $S = 5$. The transition from the paramagnetic phase to the ferromagnetic one is continuous for intermediate values of β ($\beta_t < \beta < \beta_c$) but becomes discontinuous when $\beta > \beta_t$. The transition point β_t ($\bullet$) generalizes the tricritical point of the RFIM [14] ($S = 2$).

The condition $\Lambda = 1$ on the largest eigenvalue Λ of $\beta\epsilon\mathcal{J}$ reproduces the second order transition line. The line $\Lambda = 1$ continues beyond the tricritical point and it marks the border of the region where the paramagnetic solution $\hat{m} = 0$ is unstable (dotted line in Fig. 3). Below the lower branch of the $\Lambda = 1$ line the paramagnetic solution is locally stable but it is not the most probable. Indeed the polarized state $\hat{m}^*$ which is the non-trivial solution of Eq. (10) has a lower free energy $f(\hat{m}^*) < f(0)$. Still in numerical simulation the state $\hat{m} = 0$ can persist for a very long time in this region. The polarized solution $\hat{m}^*$ becomes metastable and then disappears to the left of the transition line in Fig. 3.

With respect to the dependence on S of the phase diagram, we observe that at $\beta \to \infty$ the phase transition takes place at $\epsilon = 2/3$ independent of S. At the other ex-

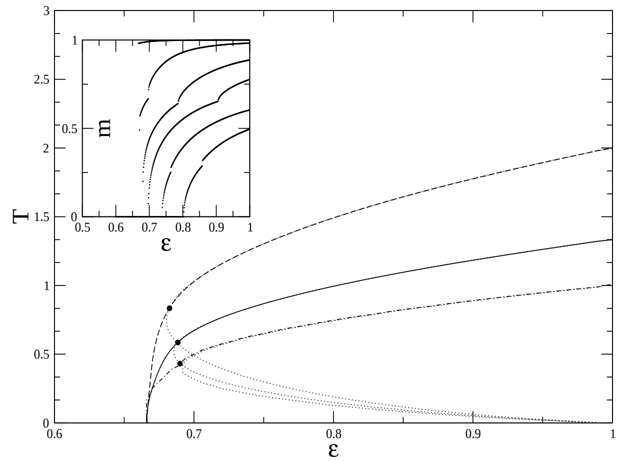

FIG. 3: Phase diagram for $S = 2,\ 3$ and 5 (dot-dashed, straight and dashed lines). Dotted lines mark the region where the $\hat{m} = 0$ phase is unstable. These meet the lines across which the transition takes place, at the tricritical point ($\bullet$). Inset: magnetization for $1/\beta = 0.25, 0.5, \ldots, 1.75$ and $S = 5$ as a function of ϵ.

treme, for $\epsilon = 1$ we find that [19] $\mathcal{J} = \mathcal{G}^{-1}$. The largest eigenvalue of $\beta\mathcal{J}$ is thus [20] $\Lambda = \beta\frac{S+1}{3}$ and the condition $\Lambda = 1$ implies that

$$\beta_c(\epsilon = 1) = \frac{3}{S+1} \tag{14}$$

Hence as S increases the region where the polarized phase is stable becomes larger and larger. In other words it becomes more and more easy for a population of agents who influence each other to become polarized on the same opinion. This is somewhat at odd with naïve expectation, because as S increases the complexity of the choice problem also increases and reaching consensus becomes more difficult. Indeed the probability $P(S)$ to find consensus on S choices in a random population drops very rapidly to zero as S increases. Nevertheless, the effects of interaction toward conformism becomes stronger. We attribute this to the fact that for large S the fraction of allowed spin configurations $\hat{v} \in \mathcal{R}$ is greatly reduced, thus inducing a strong interaction among the different spin components. This results in the fact that ordering becomes easier and easier when S increases.

B. Unconstrained case

Here the constraint $\hat{v}_i \in \mathcal{R}$ is not imposed, while we keep $\hat{\Delta} \in \mathcal{R}$. This means that an agent can be influenced by other agents' preferences to the point of picking an intransitive preference. In this case all the traces over the $\hat{v}_i$ in the above equations can be computed componentwise, independently, as in a multi-component random field Ising model. A direct computation of the matrix $J^{ab,cd}$ is possible, and yields

$$J^{ab,cd} = \delta_{ac}\delta_{bd}\left[1 - \tanh^2(\beta(1-\epsilon)\right]. \tag{15}$$

Notice that, for any β and ϵ, the maximum eigenvalue $\Lambda = \beta\epsilon[1-\tanh^2(\beta(1-\epsilon)]$ of the matrix $\beta\epsilon\mathcal{J}$ is independent of S and it coincides with that of the RFIM ($S=2$). Hence the phase diagram is that of the RFIM for all $S \geq 2$. The different spin components behave independently. The correlation induced by the constraints on the *a-priori* preferences $\Delta - i \in \mathcal{R}$ does not influence the thermodynamics properties. Note that for $\epsilon \to 1$ the condition $\Lambda = 1$ implies $\beta = 1$ and for $\beta \to \infty$ the phase transition takes place at $\epsilon = 2/3$, independent of S.

V. $P(S)$ WITH INTERACTING VOTERS

The main result of the previous section, that is, the fact that ordering becomes easier as S increases when rational voting behavior is assumed for each agent, has interesting effects on the probability of finding a transitive majority. To investigate this, we analyze the probability $P_{\beta,\epsilon}(S)$ of a transitive majority in an interacting population. The calculation is a generalization of the one presented for the non-interacting population. Let

$$\hat{z} = \frac{1}{\sqrt{N}} \sum_i^N \hat{v}_i.$$

We want to compute, at a fixed ϵ and β, the probability distribution of $\hat{z}$. Keeping fixed the realization of the disorder $\hat{\Delta}_i$, this is given by

$$P\left(\hat{z}|\{\hat{\Delta}_i\}\right) = \mathcal{N}\mathrm{Tr}_{\hat{v}_i} e^{-\beta\mathcal{H}\hat{v}_i}\delta\left(\hat{z} - \frac{1}{\sqrt{N}}\sum_i^N \hat{v}_i\right)$$
$$= \mathcal{N} e^{\frac{\beta\epsilon}{2}\hat{z}\cdot\hat{z}} \int d\hat{\lambda} e^{i\hat{\lambda}\cdot\hat{z}} \prod_{i=1}^N \mathrm{Tr}_{\hat{v}_i} e^{[\beta(1-\epsilon)\hat{\Delta}_i - i\hat{\lambda}/\sqrt{N}]\cdot\hat{v}_i}$$

now the term $\hat{\lambda}/\sqrt{N}$ is small compared to the other one and we can expand it

$$\begin{aligned}
&\mathrm{Tr}_{\hat{v}_i} e^{[\beta(1-\epsilon)\hat{\Delta}_i - i\hat{\lambda}/\sqrt{N}]\cdot\hat{v}_i} = \\
&= \mathrm{Tr}_{\hat{v}_i} e^{\beta(1-\epsilon)\hat{\Delta}_i\cdot\hat{v}_i}\left[1 - \frac{i}{\sqrt{N}}\hat{\lambda}\cdot\hat{v}_i - \frac{1}{2N}(\hat{\lambda}\cdot\hat{v}_i)^2 + \dots\right] \\
&= \mathrm{Tr}_{\hat{v}_i} e^{\beta(1-\epsilon)\hat{\Delta}_i\cdot\hat{v}_i}\left[1 - \frac{i}{\sqrt{N}}\hat{\lambda}\cdot\left\langle\hat{v}|\hat{\Delta}_i\right\rangle - \right. \\
&- \left.\frac{1}{2N}\sum_{ab,cd}\lambda^{ab}\lambda^{cd}\left\langle\hat{v}^{ab}\hat{v}^{cd}|\hat{\Delta}_i\right\rangle + \dots\right]
\end{aligned}$$

where, again, averages over the $\hat{v}$ are taken with the distribution (13). The factor $Z_i = \mathrm{Tr}_{\hat{v}_i} e^{\beta(1-\epsilon)\hat{\Delta}_i\cdot\hat{v}_i}$ can be absorbed in the normalization constant, so that if we re-exponentiate the terms, we find

$$\begin{aligned}
&\mathrm{Tr}_{\hat{v}_i} e^{[\beta(1-\epsilon)\hat{\Delta}_i - i\hat{\lambda}/\sqrt{N}]\cdot\hat{v}_i} \cong \\
&\cong Z_i e^{-\frac{i}{\sqrt{N}}\hat{\lambda}\cdot\left\langle\hat{v}|\hat{\Delta}_i\right\rangle - \frac{1}{2N}\sum_{ab,cd}\lambda^{ab}\lambda^{cd}\mathcal{J}^{ab,cd}}
\end{aligned}$$

This gives

$$\begin{aligned}
P\left(\hat{z}|\{\hat{\Delta}_i\}\right) &= \mathcal{N}' e^{\frac{\beta\epsilon}{2}\hat{z}\cdot\hat{z}} \int d\hat{\lambda} e^{i\hat{\lambda}\cdot(\hat{z}-\hat{y}) - \frac{1}{2}\hat{\lambda}\cdot\mathcal{J}\cdot\hat{\lambda}} \\
&= \mathcal{N}\text{"} e^{\frac{\beta\epsilon}{2}\hat{z}\cdot\hat{z} - \frac{1}{2}(\hat{z}-\hat{y})\cdot\mathcal{J}^{-1}\cdot(\hat{z}-\hat{y})} \\
&= \mathcal{N}\text{"} e^{-\frac{1}{2}\hat{z}\cdot[\mathcal{J}^{-1}-\beta\epsilon\mathcal{I}]\cdot\hat{z} + \hat{z}\cdot\mathcal{J}^{-1}\cdot\hat{y} - \frac{1}{2}\hat{y}\cdot\mathcal{J}^{-1}\cdot\hat{y}}
\end{aligned}$$

$$\text{where} \qquad \hat{y} = \frac{1}{\sqrt{N}}\sum_i^N \left\langle\hat{v}|\hat{\Delta}_i\right\rangle$$

and $\mathcal{J}$ given by Eq. (12). Now one needs to take the average over $P(\hat{y})$. In general this is a Gaussian distribution

$$P(\hat{y}) \propto e^{-\frac{1}{2}\hat{y}\cdot\mathcal{A}\cdot\hat{y}} \tag{16}$$

and, considering the $\hat{y}$ dependence of the normalization $\mathcal{N}$"

$$\mathcal{N}\text{"} \propto e^{\frac{1}{2}\hat{y}\cdot\mathcal{J}^{-1}\cdot\hat{y} - \frac{1}{2}\hat{y}\cdot\frac{1}{\mathcal{J}-\beta\epsilon\mathcal{J}^2}\cdot\hat{y}}$$

we get

$$P(\hat{z}) \propto e^{-\frac{1}{2}\hat{z}\cdot\mathcal{K}\cdot\hat{z}} \tag{17}$$

where

$$\mathcal{K} = \mathcal{J}^{-1} - \beta\epsilon\mathcal{I} - \frac{1}{\mathcal{J}\mathcal{A}\mathcal{J} + \frac{1}{\mathcal{J}^{-1}-\beta\epsilon\mathcal{I}}} \tag{18}$$

As before, this probability can be computed to the desired level of accuracy with the Montecarlo method.

A. Constrained case

When $\hat{v}_i \in \mathcal{R}$ we have

$$\left\{\mathcal{A}^{-1}\right\}^{ab,cd} = \left\langle\langle v^{ab}|\Delta\rangle\langle v^{cd}|\Delta\rangle\right\rangle_\Delta . \tag{19}$$

Fig. 1 ($\diamond$) shows that the resulting $P_{\beta,\epsilon}(S)$ may exhibit a non-monotonic behavior with S: first it decreases as $P(S)$ and then, as the point (β,ϵ) approaches the phase transition line it starts increasing. If $\epsilon > 2/3$, there is a value S^* beyond which the system enters in the polarized phase and $P_{\beta,\epsilon}(S) = 1 \; \forall S \geq S^*$.

B. Unconstrained case

In this case $\left\langle\hat{v}|\hat{\Delta}_i\right\rangle = t\hat{\Delta}_i$ where we introduce the shorthand $t = \tanh[\beta(1-\epsilon)]$. Then $\mathcal{A} = \mathcal{G}/t^2$ or

$$P(\hat{y}) \propto e^{-\frac{1}{2t^2}\hat{y}\cdot\mathcal{G}\cdot\hat{y}} \tag{20}$$

in addition

$$\mathcal{J} = (1-t^2)\mathcal{I} \tag{21}$$

hence setting $f = 1 - \beta\epsilon(1-t^2)$

$$\mathcal{K} = \left[\frac{1}{1-t^2} - \beta\epsilon\right]\mathcal{I} - \frac{t^2}{1-t^2}\frac{f}{t^2\mathcal{I} + f(1-t^2)\mathcal{G}}. \tag{22}$$

The behavior of the probability can be understood in some interesting limits. For $\beta \to \infty$ we get

$$\mathcal{K} \simeq \mathcal{G} + O(1-t^2)$$

which simply states that as the temperature goes to zero the probability reduces to that of the constrained case, as it should. Note that $\mathcal{K} \to \mathcal{G}$ also as we approach the critical line where $1 - \beta\epsilon(1-t^2) \to 0$.

Instead for $\epsilon \to 0$ we have

$$\mathcal{K} \to \frac{1}{1-t^2}\left[\mathcal{I} - \frac{1}{\mathcal{I} + (t^{-2}-1)\mathcal{G}}\right]$$

The high T limit $\beta \to 0$ reads

$$\mathcal{K} \simeq \mathcal{I} - \beta^2\mathcal{G}^{-1} + \dots$$

that is, since the matrix $\mathcal{K}$ is diagonal the probability of finding a transitive majority drops to the trivial one, namely $S!2^{-S(S-1)/2}$. So without the constraint of rational voting the probability of a transitive outcome can be greatly reduced. Again, Monte Carlo simulations are shown in Fig.2 ($*$). Note the marked decrease of the probability of finding a transitive majority with respect to the constrained and to the non-interacting case.

VI. CONCLUSIONS

In conclusion we have studied the properties of pairwise majority voting in random populations. We have computed the probability that pairwise majority is transitive when there is no interaction and found that it decreases rapidly with S, even though less rapidly than one would naively guess. Then we have shown that the properties of pairwise majority in a random interacting population are related to the properties of a multi-component RFIM, whith a constraint on the components which reflects the transitivity of individual preferences. This model can be solved exactly and features a ferromagnetic phase where the population reaches a consensus (i.e. a transitive majority) with probability one. As to the dependance on the number of voters, we find that the ferromagnetic phase gets larger and larger as S increases, meaning that consensus is reached more easily when the complexity of the problem (i. e. the number of alternatives) is large enough.

With respect to the case when rational voting behavior is not imposed, we note the strikingly different effect that intereaction can have, dependant on how this interaction is introduced. In fact, if we impose a transitive voting behavior, the probability to find a transitive majority is increased, while relaxing this constraint can result in a decrease of this probability.

[1] T. M. Liggett, Interacting Particle Systems (Springer, New York, 1985).

[2] S. Galam, Physica A **238**, 66 (1997).

[3] L. F. C. Pereira and F. G. B. Moreira, cond-mat/0408706.

[4] S. N. Durlauf, in *The economy as an evolving complex system II*, W. B. Arthur, S. N. Durlauf and D. A. Lane (eds.) (Addison-Wesley, 1997).

[5] T. Schelling *Micromotives and macrobehavior*, (W. W. Norton, 1971)

[6] Condorcet *Essai sur l'application de l'analyse a la probabilt des decisions rendues a la pluralit des voix* 1785

[7] D. A. Meyer and T. A. Brown, Phys. Rev. Lett. **81**, 1718 (1998).

[8] P. Dasgupta and E. Maskin, Scientific American **290** (3), 92-97 (2004)

[9] K.J. Arrow, *Social Choice and Individual Values* (Wiley, 1951).

[10] G.T. Gilbaud, Economie Appliquee **5**, 501-584 (1952).

[11] R. Niemi and H. Weisberg, Behav. Sci. **13**, 317-323 (1968)

[12] W.V. Gehrlein Public Choice, **66**, 253 (1990).

[13] T. Schneider and E. Pytte, Phys. Rev. B **15**, 1519-1522 (1977)

[14] A. Aharony, Phys. Rev. B **18**, 3318-3327

[15] H. P. Young, in *Social Dynamics*, H. P. Young and S. N. Durlauf (eds.) (Brooking Institution, 2001)

[16] M. Kandori, G. Mailath, and R. Rob, Econometrica **61**, 29-56 (1993).

[17] S.N. Durlauf, Proc. Natl. Acad. Sci. **96**, 10582 (1999).

[18] A similar picture was found with an adaptive dynamics where agent learn about their best choice over time.

[19] With $\epsilon = 1$, $J^{ab,cd} = \frac{1}{S!}\sum_{\hat{v}\in\mathcal{R}} v^{ab}v^{c,d}$. So $J^{ab,ab} = 1$ and $J^{ab,cd} = 0$ by symmetry if $a \neq c, d$ and $b \neq c, d$. Furthermore $J^{ab,ad}$ depends only on the relative ordering between a, b and d in the permutation $\hat{v}$. The permutations where a is between b and d, which are 1/3, give $v^{ab}v^{ad} = -1$, whereas the remaining $\hat{v}$ give $v^{ab}v^{ad} = 1$. Hence $J^{ab,ad} = \frac{\beta}{3}$. Likewise we find $J^{ab,cb} = -J^{ab,ca} = -J^{ab,bd} = \frac{\beta}{3}$.

[20] The matrix $\mathcal{G}$ has $S-1$ eigenvectors of the form $z_{1,k}^{ab} = \delta_{a,k-1}\mathrm{sign}(b-k+1)$ with $k = 2, \dots, S$, and eigenvalue $\lambda = 3/(S+1)$. This can be verified by direct calculation. Note that the vectors z_k^{ab} are not orthogonal, but are linearly independent. Direct substitution shows that all

vectors of the form $z_{j,k}^{ab} = \delta_{a,1}(\delta_{b,j} - \delta_{b,k}) + \delta_{a,j}\delta_{b,k}$ are also eigenvectors of $\mathcal{G}$ for $1 < j < k$, with eigenvalue $\lambda = 3$. Then $\lambda = 3$ has degeneracy $(S-1)(S-2)/2$. The set of $S(S-1)/2$ linearly independent vectors $z_{j,k}^{ab}$ with $1 \leq j < k \leq S$ allows us to build a complete orthonormal basis of eigenvectors and to compute $\det \mathcal{G}$.